Title Cloud-computing Approach to Sustainable Media Archives

Authors Erik Mannens (UGent-MMLab), Sam Coppens (UGent-MMLab), Rik Van de Walle (UGent-MMLab), Laurence Hauttekeete (UGent-MICT), Robbie De Sutter (VRT)

In: Proceedings of the 2010 FIAT/IFTA World Conference - Making it Real, Keeping it Real (FIAT/IFTA 2010), 2010, October

Optional: http://www.conferencepartners.ie/fiat_ifta/text2.html

# Cloud-computing Approach to sustainable Media Archives

Erik Mannens, Sam Coppens, Rik Van de Walle[1], Laurence Hauttekeete[2], Robbie De Sutter[3]

As the current multimedia assets of each institution –be it a content producer, an archive, a cultural institution, library, or museum– are governed in a 'geographically distributed' way by nature and the internet is currently by far the easiest 'physical medium' to use for connecting those data islands. It is clear that also 'cloud computing' should at least be considered the new paradigm to describe the production, alteration, consumption and delivery model for IT services based on the Internet, as it typically involves the provision of dynamically scalable and often virtualized resources as a service over the Internet. It is indeed a byproduct and consequence of the ease-of-access to remote computing sites provided by the nowadays omnipresent Internet.

As such the Flemish Archipel project initiates the digital long-term preservation of cross-sectoral cultural, media heritage in Flanders and researches the problems encountered with distributed digital long-term preservation of multimedia. To overcome all the individual risks, the data needs to be described on different levels, i.e. from bitstream level to the intellectual entity level, thereby securing all the rights and requirements of the individual cultural institutions. By providing a three-layered semantic metadata model, not only the metadata are stored, but also the semantics of the metadata are stored for the long term. The top layer (for initial exchange) is a representation of the descriptive metadata in Dublin Core, the intermediate layer (the cross-sectoral refinement) is a more detailed description of the specific sectoral standards involved (MARC, CDWA, P/Meta, EAD, Spectrum, ISAD-G), and the bottom layer (for long-term preservation purposes), is modeled via PREMIS.

This model must be used in combination with preservation strategies, which ensure the accessibility of the digital objects for the future. These preservation strategies consist of several workflows for each file format, accepted by the preservation platform. These workflows put the digital object on a trajectory of certain actions, like validation, virus checking, normalisation, ingest, migration, emulation, etc., to ensure the future access to the digital object. These actions can be modelled perfectly within a distributed service model as PREMIS events. For this reason, we made a binding of our workflow engine, which executes the preservation strategies, to our developed metadata model. This way, the workflow engine can be used in any digital repository turning it into a distributed digital long-term archive, assuring the digital preservation.

As such, sustainability is key in Flanders effort of converging their valuable Archiving, Libraries, Museum, and Broadcasting assets into a distributed test bed where OAIS-compliant submission and dissemination modules are developed on top of the decentralized Fedora framework while being compatible with other European initiatives. This project delivered the Flemish consortium both technical, organizational, and strategic innovative insights in the archiving challenges at hand.

Our findings from this project form the basis of our view on future distributed media archive systems described in this paper.

[1] Ghent University – IBBT, Multimedia Lab, G. Crommenlaan 8, 9050 Ghent, Belgium
[2] Ghent University – IBBT, MICT, Korte Meer 7-9-11, 9000 Ghent, Belgium
[3] VRT, August Reyerslaan 52, 1043 Brussel, Belgium